(12) UK Patent Application (19) GB (11) 2 339 905 (13) A

(43) Date of A Publication 09.02.2000

(21) Application No 9914723.3

(22) Date of Filing 23.06.1999

(30) Priority Data
(31) 19828051 (32) 24.06.1998 (33) DE

(71) Applicant(s)
Bruker Daltonik GmbH
(Incorporated in the Federal Republic of Germany)
Fahrenheitstraße 4, D-28359 Bremen,
Federal Republic of Germany

(72) Inventor(s)
Thomas Bonk
Andreas Humeney
Cord-Michael Becker
Magnus Von Knebel Doeberitz
Jochen Franzen

(74) Agent and/or Address for Service
W H Beck, Greener & Co
7 Stone Buildings, Lincoln's Inn, LONDON, WC2A 3SZ,
United Kingdom

(51) INT CL[7]
C12Q 1/68

(52) UK CL (Edition R )
G1B BAC B100 B200 B202 B223

(56) Documents Cited
WO 98/20166 A2 WO 97/33000 A1
Rapid Communications in Mass Spectrometry 11(15) 1997 Wada, Y. et al. pages 1657-1660. Rapid Communications in Mass Spectrometry 11(10) 1997 Jannavi, R-S. et al. pages 1144-1150.

(58) Field of Search
UK CL (Edition Q ) G1B BAC
INT CL[6] C12Q 1/68
ONLINE: WPI, EPODOC, CAS-ONLINE.

(54) Abstract Title
**Use of mass-specrometry for detection of mutations**

(57) The present invention relates to a method for the mass spectrometric recognition of polymorphisms and mutations in a nucleic acid comprising:

- providing a sample of double-stranded nucleic acid segments, preferably from one full gene or exon, and preferably obtained through PCR amplification.
- adding restriction enzymes that will digest said nucleic acids into a mixture of double-stranded fragments of 10 to 40 bases in length.
- ionizing the resulting mixture.
- determining the molecular weights of the digest fragments by mass spectrometry.
- determining mutative changes or variations in the digest fragments by comparing the
- molecular weights thereof with those of a reference DNA digested with the same set of endonucleases.

Also claimed are kits for performing the above method.

GB 2 339 905 A

# Measurement Method for Polymorphisms and Mutations in Nucleic Acids

The invention relates to a method and a chemical kit for rapid, mass-spectrometric screening measurement of polymorphisms and mutations in nucleic acids, preferably in genes or gene segments, and the associated preparation of samples from amplified DNA.

Desoxyribonucleic acid (DNA) consists of two complementary chains of four nucleotides (adenine, cytosine, guanine and thymidine), designated by the letters A, C, G and T, the sequence of which - at least in the coding parts of the DNA - encodes the formation of proteins by the genetic code. The two complementary chains of DNA are most commonly arranged in the shape of a double helix, whereby two complementary nucleotides each are joined together between the bases via two (adenine = thymidine) or three hydrogen bridges (cytosine ≡ guanine).

The genetic code is an encryption of the sequence of amino acids within the proteins thus formed. However the approximately 120,000 proteins in a human being, with their multiple functions, are encoded by only about three percent of human DNA, comprised in total of about 3.5 billion base pairs; the remainder of the DNA is noncoding and only partly has regulatory functions (promoter regions, enhancers, silencers). With the exception of splicing variants, each protein is encoded by one gene.

The sections of a gene that code for protein sequences ("exons") are generally interrupted by larger or smaller islands of noncoding DNA (so-called "introns"). After transcription of the DNA sequences into ribonucleic acid (RNA), these introns are excised by a so-called splicing procedure, whereby this process is controlled by those DNA segments of introns which border on affiliated exons (so-called "splicing signals"). The DNA sequences within the introns frequently demonstrate extreme variation between various people or various individuals of a species, since these types of mutations have no influence on the

protein structure, as long as the splicing process is not inhibited, and thus are subject to only minimal evolutionary selection. Within the exonic regions, the DNA sequence does not vary so extremely, because individuals that have mutations with a negative or even lethal effect on proteins formed during the translation are culled out. In spite of this, a more or less extreme degree of variability ("polymorphism") also prevails in genes ("genotype"), which

partly has no functional effect at all because of the redundancy of the genetic code,

is partly expressed in the external appearance of individuals (the "phenotype"),

or partly influences only the metabolism or other bodily functions in a manner not externally apparent.

The basis for most DNA analysis methods is its amplification by the selectively operating PCR ("polymerase chain reaction"), a simple amplification method that can be carried out in a test tube for specifically selectable DNA sections. This method was developed in 1983 by K.B. Mullis (who received the Nobel Prize for this in 1993) and after the introduction of thermally stable DNA polymerases, it became generally accepted in genetic laboratories. Similar enzymatic replications also exist today for RNA, whereby PCR is preceded, for example, by a reverse transcription step from RNA to DNA performed *in vitro*.

PCR is the specific amplification of a defined section of double-stranded DNA ("dsDNA"). A DNA segment is selected by a pair of so-called primers, two single-stranded DNA segments ("ssDNA") of about 20 bases length, each sequentially homologous to the end pieces of the selected DNA section. These primers, in a simplified description, attach ("hybridize") to both sides (the future ends) of the required DNA piece. Amplification is conducted with an enzyme called DNA polymerase in a simple thermal cycle which will not be discussed here in any greater detail.

Currently, the lengths of PCR-amplified DNA segments are mostly measured by means of the process of electrophoresis in agarose or polyacrylamide gels, which is slow and not capable of being completely automated . Here, the molecular weights are measured by ion mobilities in the gel under the influence of an electrical field. This pioneering method is based upon lengths of DNA segments of about 50 to 5,000 base pairs; its precision in measuring mobility, and thus for measuring the length of the sequence, is very limited. Between two DNA segments of differing lengths with approx. 500 base pairs, differences of one base pair can be just distinguished; the recognition of point mutations ("substitution of a base pair") is clearly impossible. In longer segments, sequence insertions and missing sequence sections ("deletions") can also no longer be recognized.

Experts see no future in this analysis method, particularly because automation, often attempted, continues to remain impossible due to frequently occurring artefacts of different types, and is thus very labor and time consuming, and also quite expensive because of the relatively high consumption of expensive reagents. This procedure was a pioneering method which however increasingly proved to be a bottleneck for further propagation of genetic analysis.

It can be expected that various mass-spectrometric measurement methods will be qualified to determine the molecular weight of DNA fragments with much higher rates of measurement, much greater reliability and much improved measurement precision than the determination of mobility by means of electrophoretic or chromatographic methods. Ionization of DNA segments using the known methods of electrospray (ESI) and matrix-assisted laser desorption and ionization (MALDI) are also applied for mass-spectrometric measurements. High-frequency quadrupole ion traps, ion-cyclotron resonance spectrometers, or time-of-flight spectrometers may particularly be used as mass spectrometers for these measurements. An especially favorable combination is MALDI and a time-of-flight mass spectrometer (TOF-MS)

The current rapid progress in the MALDI technique is leading to a high degree of automation for sample ionization and to short analysis times per sample. In a time-of-flight mass spectrometer, ions fly about $10^7$ times faster than they move electrophoretically through gel. Even if about 10 to 100 mass spectra are necessary for a signal with a good signal-to-noise ratio, the mass-spectrometric method is much faster by far. In this way, very precise molecular weight determinations for analyte molecules of up to about 15,000 atomic mass units (about 40 bases ssDNA) in size are possible in quantities of several femtomole over measuring periods of a few seconds. Thousands of samples can be applied to one sample support. Automation and a high density of the samples open up the possibility of processing several tens of thousands of samples per day with mass-spectrometric analysis.

Ionization by means of matrix-assisted laser desorption (MALDI) is an ionization method for macromolecular analyte substances on surfaces. The analyte substances are applied in a solution to a surface with suitable matrix substances of a lower molecular weight than the analyte molecules. There they are dried and irradiated with a laser pulse of a few nanoseconds duration. A minimal amount of matrix substance vaporizes, a few molecules of which as ions. The very dilute analyte substance in the matrix, the molecules of which are distributed throughout the dilution, are also vaporized, even if their vapor pressure would not normally suffice for vaporization. The relatively small ions of the matrix substance react with the large molecules of the analyte substance, so that the analyte substance consequently remains behind in the form of ions for energetic reasons due to proton transfer. The double-stranded DNA (dsDNA) is denatured into single-stranded DNA (ssDNA) within the MALDI process.

Mass spectrometers using the MALDI or ESI technique are already being utilized for sequencing of DNA according to the Sanger scheme while making use of PCR methods. In reference to this see PCT/US94/00193 or Kirpekar et al. 1998, Nucleic Acids Research 26(11), 2554-2559.

The mass-spectrometric measurement of DNA segment masses is also subject to restrictions. In contrast to gel-electrophoretic methods, mass spectrometry can presently only measure very small DNA pieces. The limit currently lies at about 60 to 120 bases. Because of the negatively charged phosphoric acid groups of the backbone of the DNA it is necessary to multiply protonate the DNA fragment to become a positive ion. For MALDI, this requires very special matrix substances. Only a few matrices, most preferably 3-hydroxypicolinic acid (HPA), can be used here. Larger DNA ions suffer as well from fragmentation as from formation of adducts with other positive ions around, such like sodium or potassium ions. Fragmentation and adduct formation both broaden the mass peaks, limiting the accuracy of mass determination at higher DNA segment masses.

On the other hand, a precision of mass determination which extends far beyond that of electrophoresis can be achieved in the lower mass range provided cleaning of the DNA pieces is good. The molecular weight of DNA segments up to 20 bases in length can be determined to a tenth of an atomic mass unit, up to 30 bases in length to a half mass unit, and up to 40 bases in length to about five mass units. Since the minimal mass difference of two bases is nine mass units (adenine and thymidine), substitution of a base by another can be recognized with certainty. It is therefore essential for precise mass determination to operate with the lowest possible molecular weights for DNA chains. To this end, reflector time-of-flight mass spectrometers are particularly useful. On the other hand, it is absolutely essential to remove any salts containing non-degradable cations, such as sodium or potassium, to avoid adduct formation.

One of the objects of this invention is a method for simultaneous and rapid detection of one or several known polymorphous or mutative changes within an specific sequence of genomic or mitochondrial DNA from an organism, preferably sequences of a gene ("gene screening"), or a DNA derived from RNA through reverse transcription, in contrast to sequences of a standard DNA often designated as "wild type". For these sequence changes, it

may concern a base substitution ("point mutation"), insertions of one or several bases ("insertion mutation") or the lack of one or several bases ("deletion mutation").

As mentioned, one gene generally encodes one protein each, which furthermore has a specific function in human, animal or plant bodies, or for bacteria or viruses. Proteins deviating from the standard may - without displaying any particular malfunctioning - lead to a changed phenotype of an individual. It may however also lead to changed reactions of the body to internal or external influences, for example to changed reactions to drugs. Genotype-dependent medications will play an important role in therapeutic applications in the future. In order to clearly characterize mutations, a DNA sequence in which a mutation is suspected must be sequenced. To then find an identified mutation in another individual, an existing analysis can be used: renewed sequencing of the corresponding DNA segment. In practice, this would mean that detection of a known mutation in a test subject would generate the same costs as the original characterization. For sequencing, various types of gel electrophoresis are used which nevertheless, as mentioned above, are slow and not completely automatable - and are consequently expensive.

For this reason, alternative and less expensive techniques have been developed for detecting the presence of known mutations. For example, appropriate DNA sequences can be combined by means of surface fixation on a DNA chip for simultaneous identification of many mutations. Their hybridization or non-hybridization with applied genetic material can be used by comparison with a standard DNA pattern for simultaneous determination of a large number of various mutations. Thus chips are known with 64,000 systematically varied and fixed sequences. This DNA chip technology nevertheless has several important disadvantages. On the one hand, manufacture of the DNA chips is quite expensive and the chips are not reusable. Additionally, this method - like all hybridization methods with relatively small sequence anchors - cannot be validated medically due to its inherent uncertainty. It

represents a relatively good screening method, but cannot yet be used for safe diagnosis of a specific, defined illness.

For the exclusive diagnosis of known point, insertion, or deletion mutations, a new method has recently become known which utilizes MALDI mass spectrometry (Little, D.P., Braun, A, Darnhofer-Demar, B., Frilling, A., Li, Y., McIver, R.T. and Köster, H.; Detection of RET proto-oncogene codon 634 mutations using mass spectrometry. J. Mol. Med. 75, 745-750, 1997). The primer (a DNA chain functioning as a recognition sequence) is synthesized here in such a way that it attaches ("hybridizes") itself in the immediate vicinity of a known point mutation on the template strand. Between the position of this mutation and the 3' end of the primer (the primer is extended at this end), the sequence of the template strand must be comprised of a maximum of three of the four nucleobases. At the mutation position, a further base appears for the first time. Using a polymerase and a special set of deoxynucleotide triphosphates (a maximum of three complementary ones which occur up to the point mutation position) and a dideoxynucleotide triphosphate (with a base that is complementary to the potential mutation), the primer is extended by duplication. The dideoxynucleotide triphosphate terminates the chain extension. Depending on the presence or absence of a mutation, the polymerase reaction is terminated at the point mutation position or it terminates just at the next corresponding base beyond the potential mutation point. This method, which however also includes a fixation of the primer to a surface, has been designated as "PROBE" by the authors. This method, specially developed for mass spectrometric analysis, is restricted to the relocation of a very precisely known mutation. It can neither be used as a screening method for unknown mutations nor simultaneously analyze larger numbers of potential mutation points.

On the other hand, one of the objects of the present invention is a simple method to find new and hitherto unknown mutations in a given gene.

In genetic biochemistry, a method for analysis of restriction-fragment-length polymorphism (RFLP) is known that can detect a certain number of unknown mutations. The RFLP method consists or subjecting large DNA pieces from an enzymatic splitting ("digest") through one or several restriction enzymes, so-called "restriction endonucleases". These restriction enzymes have a particular detection mechanism for a fixed sequence of four to eight base pairs and cut the DNA at a defined point. The resulting DNA segments are then subjected to gel electrophoresis which determines the length of individual segments. Since gel electrophoresis is not capable of detecting point mutations or, for larger fragments, changes in length caused by shorter insertions or deletions, only those length changes are recognized here (always compared to the pattern of standard DNA) that are produced by such mutations which prevent cutting by the enzyme or insert an additional cutting point. Such mutations are usually found in the more variable introns, much more rarely in exons.

The so-called SSCP method ("single strand confirmation polymorphism") represents an alternate method for detection of DNA polymorphisms by which DNA fragments generated by PCR from about 100 base pairs after denaturing (transformation of double-stranded DNA into single-stranded DNA) are subjected to polyacrylamide gel electrophoresis at generally four different temperatures. Here, single strands can take on a changed three-dimensional structure ("confirmation") based on point mutations, which can be expressed in differing mobilities compared to wild type DNA. In this method, substitutions of individual nucleotide positions between two DNA fragments with otherwise identical sequences are only detected in approximately 70% of cases. Since this method and its variants (for example, those based on chromatography) are additionally associated with high personnel and time expense, it has not been considered for a mass screening.

Therefore there is still a need for methods that can quickly, safely and inexpensively detect mutations of individual genes, whereby a higher degree of multiplexability is desirable for

simultaneous analysis of many potential mutation points, which however must not inhibit its ability to be validated. It would be favorable if this method could also be used inexpensively for reliable analysis of larger groups of people as to variability and typing of specific genes, to detect all polymorphisms or mutations of a gene or of segments of a gene, including the enormous quantity of presently unknown mutations.

A mass spectrometric method to find mutations has become known from WO 97/33 000. The target nucleic acid is fragmented to obtain a set of nonrandom length fragments (NLFs) in single-stranded form, either chemically or enzymatically, and the fragment masses are determined by mass spectrometry. This can be achieved, among other methods, by restriction endonucleases. The mutations can be detected by comparison of precise masses with those of the NLFs of a wild type DNA. Another method is to generate double-stranded NLFs wherein the fragmenting comprises using volatile salts in a restriction buffer.

Objective of the invention

Accordingly, it would be desirable to find a mass-spectrometric method for fast, simultaneous, inexpensive and validatable measurement of all (known as well as unknown) polymorphisms and mutations in nucleic acids, particularly in genes or in gene segments, and to provide the genetic material and delivery of support material in kit form necessary for this.

The method should desirably be able to be used to

search for mutations in larger population groups ("genetic screening"),

as well as reliably relocate mutations present in individual test subjects,

detect mutations limited to individual bodily or tumor tissue ("somatic"),

search and/or detect genetic variants which, although they remain without recognizable effect on the organism, permit genetic identification of individuals or population sections ("genetic fingerprint") and

detect mutations in organisms of all types.

Both preparation as well as data acquisition and evaluation should be performed as automatically as possible with commercially manufactured pipetting robots and suitable computer software.

In a first aspect of the invention, there is provided a method for the mass spectrometric recognition of polymorphisms and mutations in a nucleic acid, comprising the following steps:
(1) providing amounts of double-stranded target segments of nucleic acid, preferably from one full or partial gene,
(2) adding a set of restriction enzymes, operating at similar buffer conditions at defined restriction points, to the DNA target segments, and digesting the target segments into a mixture of double-stranded DNA digest fragments of about 10 to 40 bases in length,
(3) removing from the mixture cations which may result in adduct formation by the ionization method chosen,
(4) determining the molecular weights of digest fragments of the mixture by mass spectrometry, and
(5) determining mutative changes or variations in the digest fragments by comparing the molecular weights of the digest fragments with those of a reference DNA, digested with the same set of endonucleases.

The invention consists of cutting the amplified DNA target segments by means of a tailored set of restriction endonucleases at defined points into small digest fragments of DNA, and to examine the mixture of digest fragments after special cleaning in a mass spectrometer for the molecular weights of the digest fragments. The deviations in molecular weights of these fragments from those of a reference DNA (a "standard DNA" or so-called "wild type") indicate all polymorphisms or mutations present in the DNA target segment investigated. The set of endonucleases is tailored to contain only such endonucleases which operate at very similar buffer conditions, and which generate a favorable set of digest fragments in a certain mass range and with no overlap of the isotopic pattern.

The invention provides DNA target segments of a gene ("exon") in sufficient quantity and subjects these target segments of double-stranded DNA to a specific digestion by a set of restriction enzymes so that double-stranded DNA digest fragments in the range of about 10 to 40 base pairs result. This mixture of DNA digest fragments is thoroughly purified from all non-decomposing salt ions and subjected to mass-spectrometric analysis of the molecular weights, preferably by MALDI, and mutations are recognized by differences in the molecular weights of the same digest fragments of standard DNA.

The DNA target segments are preferably supplied by means of amplification by a PCR method, either as single pieces or as multiples by multiplexed PCR, and thorough cleaning from all nucleotide triphosphates and primers.

If necessary, intronic sections may also be included in the procedure, for example those intronic sequences necessary for the splicing process or those regulatory DNA sections located upstream or downstream from the coding gene section.

There are at present more than 2000 restriction endonucleases known which differ, however, in optimum salt concentration and pH conditions (and price). The restriction points of the endonucleases cover nicely any possible DNA sequence, there is at least one nuclease available for a restriction at every sixth base in maximum, in most cases, the distances of possible restrictions are nearer to each other. There are endonucleases which cut straight through both strands of the DNA, others cut sense and anti sense strands at slightly different points. Restriction data like recognition sequences and operation conditions for the known endonucleases can be accessed through internet. Programs are available which calculate and mark all restriction points in a given DNA sequence.

It is part of the invention to select a set of endonucleases in such a manner that (a) DNA digest fragments are generated with about 10 to 40 bases in length, (b) no severe overlap of the resulting isotopic peak pattern results, and (c) only such endonucleases are selected which operate at similar conditions so

that they can be used in the same buffer. The endonucleases can be selected either experimentally, if the DNA sequences of the target fragments are not known, or else by an selection algorithm, e. g. by an computer program, taking into account the known conditions for the endonucleases.

For genes in which the required genomic sequences are not yet known, the RNA of those tissues in which these genes are activated ("expressed"), can be used as original material. This no longer contains the intronic sequences. The RNA can be transformed, for example, into DNA by means of a reverse transcriptase which is then replicated using the PCR method.

In contrast to the method of WO 97/33 000 producing double-stranded DNA digest fragments using volatile salts in the digest buffer, this invention relies on extensive cleaning of the DNA digest fragments from all non-decomposing salt cations in the buffer used for the enzymatic restriction.

From mass spectrometric determination of molecular weights for specific, relatively short digest fragments, the lengths of which are known from standard DNA (or another type of comparison DNA), the presence of a mutation can be safely established. The mutations need not be previously known for this method. The usually overwhelming number of digest fragments which are equal in mass to those of the standard DNA, may be used as internal mass reference peaks to improve the mass determination of the digest fragments which differ in mass.

With a relatively short gene, all exons of the gene can be subjected simultaneously to multiplexed PCR since the number of digest fragments remains relatively low within the mixture to be analyzed mass-spectrometrically. Since the masses of these digest fragments are measured in one single mass-spectrometric analysis, they should be distinguishable from one another; in this way the number of digest fragments in one analysis is limited. However, for a large gene with more than 500 base pairs, it is more practical to analyze the gene in overlapping segments rather than the entire gene in a single mass-spectrometric analysis, for example by using appropriate primers within the PCR process.

Particularly for the analysis of mutations in nucleic acids (DNA or RNA), the necessary chemicals and tools for reverse transcription, for PCR replication, for enzymatic digestion and for purification of the PCR products and digest products, are assembled in corresponding kits and can thus easily be used. The tools may consist of, for example, magnetic beads for temporary surface bonding for the purpose of washing, as well as chromatographic mini-tubes or prepared pipette tips for purification. The kits can be assembled and packed in such a way that they can be further processed by automatically functioning pipetting robots.

Particularly favorable embodiments

The method of the invention aims primarily for the encoding section of genes and it finds practically all mutations, not just the relatively rare, cut-changing mutations as in RFLP. Only exceptions are, as mentioned above, the extremely rare compensating double mutations (e.g. base rotations) within a digest fragment, because they do not change the mass, insofar as they do not create or eliminate a restriction cutting point on their part, or represent an obstacle for DNA amplification.

A suitable embodiment of the method for a short gene of a still unknown sequence with a maximum of about 500 base pairs length appears as follows: The gene is extracted in the usual manner as RNA from cells, passed through reverse transcriptase in DNA and amplified by means of PCR. To do this, only the sequences of the end pieces need be known so that the corresponding primers can be synthesized. The amplified products are purified in the usual manner to remove the residual nucleotide triphosphates and primers, and then subjected to simultaneous digestion by a first set of various restriction enzymes.

The restriction enzymes recognize a specific sequence of four to eight base pairs. A restriction enzyme with a recognition sequence of four base pairs cuts the DNA in average lengths of 256 base pairs. If ten differently cutting enzymes are used simultaneously, digest fragments of about 26 bases in length should result in statistical average. At an average length of 26

base pairs, about 20 dsDNA digest fragments of mainly about 10 to 40 base pairs in length are formed from one dsDNA piece of 500 bases in length.

These digest fragments are then purified using corresponding tools from all buffer cations, mixed with matrix, applied to a sample support, transferred into the mass spectrometer, ionized by a laser pulse and analyzed as ions in a high resolution mass spectrometer. Since the double strands are denatured to single strands in the MALDI process, about 40 signal groups of single-stranded DNA digest fragments result, which must be measured in a mass range of about 3,000 to 12,000 atomic mass units. If the length of the digest fragments is somewhat evenly distributed over this range by an optimum set of restriction enzymes, the average distance between the digest fragments of different lengths is roughly 225 atomic mass units. Since point mutations (SNPs = single nucleotide polymorphisms) show differences in lengths of 9 to 40 atomic mass units, overlaps even with expected mass changes by point mutations may easily be avoided by a suitable set of restriction endonucleases.

The signal groups each consist of the monoisotopic base peak, which may be very small, and the satellite peaks formed by the isotopes. Below about 6000 atomic mass units, the isotopes are usually separated in a good mass spectrometer. Starting with digest fragments of about 20 base pairs length, the isotopic satellite peaks in the signal group merge with the monoisotopic peak into one single signal. At mass $m$ = 12000 atomic mass units, the maximum of the isotopic pattern is at $m$ + 10 atomic mass units, if $m$ is the isotopic mass. The molecular weight of each signal group can be measured with an accuracy of better than a few atomic mass units. By comparison with the corresponding molecular weights of the digest fragments of a standard DNA, the digest fragments containing mutations can be found and even the type of mutations can be determined. Thus several types of mutations can even be distinguished in one DNA target segment, or even in a single digest fragment, as long as they mostly appear

one at a time. For more information on the mutation, the digest fragment has to be sequenced in detail.

If short digest fragments much below 10 base pairs or if overlaps of the isotopic pattern of the digest fragment mass peaks occur, the set of restriction enzymes has to be changed experimentally, and a second measurement has to be performed to see the possible improvement. This process has possibly repeated several times, until an optimum set of enzymes is found. If long digest fragments are produced, additional, more specific enzymes with longer recognition sequences can be tried.

This experimental optimization of the set of enzymes may be a lengthy procedure. But if an optimum set has been established, this set can be used over and over for the same gene target fragment to either screen a population for unknown mutations or to determine known mutations in individuals.

But fortunately we are approaching a situation where we know most of the sequences of the human genes, and many genes of animals or plants of interest. The optimum selection of restriction enzymes becomes then much easier. We can select enzymes and predict theoretically the lengths of the digest fragments produced by these enzymes. We can even write a program to find an optimum set of enzymes with no overlap of the resulting mass peak isotopic pattern and almost evenly distributed lengths of the digest fragments, choosing enzymes of most equal digest conditions, and even of lowest price.

The program may start with an standard mix of about six to ten enzymes with recognition lengths of four bases each, as described above. If digest fragments of too short a length are predicted, one of the enzymes causing the short digest fragment may be left off or exchanged. In a similar method, overlaps can be avoided. For too lengthy digest fragments, suitable enzymes with recognition sequence lengths of six to eight bases may be added to cut these digest fragments without affecting the residual digest fragments.

Since mutations are rather rare in the DNA, most of the digest fragment masses are not altered. The unaltered masses can be used

as mass reference peaks in the resulting mass spectrum. This can be done easily by fitting the curve of all the expected masses to that of the found masses, resulting in a smooth curve for all unaltered masses. The digest fragments altered in length then clearly deviate from this smooth curve. The smooth part of the curve represents the "mass calibration curve" for the mass spectrometer.

Instead of a comparison with a standard DNA ("wild type DNA") as the comparison reference, the DNA digest fragment masses of a diseased subject may be compared with those of a closely related healthy subject (a subject of the same family) as a comparison reference to concentrate on those mutations which may be correlated with the disease and somewhat suppress mutations between unrelated subjects of this species.

Statistical analysis of a larger population group shows the variability of individual gene digest fragment masses and thus the correspondent mutations. The effect of individual mutations on certain diseases can be calculated through coupling analyses, comparing mutations in healthy and diseased populations.

For an individual test subject, where there is suspicion of a mutation in a gene, the same method can be applied for the determination of a known or even unknown mutation. If the gene with its sequence is known, it is possible to proceed directly from the DNA. If the gene is short enough, the exons of the gene can be simultaneously replicated by means of multiplexed PCR.

To determine the molecular weights of DNA digest fragments, it is a preferred method to use MALDI ionization with analysis in a time-of-flight mass spectrometer. A time-of-flight mass spectrometer with delayed acceleration and energy-focusing ion reflector is particularly advantageous, which leads to good mass determination via high mass resolution. However, use of other mass spectrometers is also possible. For instance, highest mass accuracy will be achieved in ion cyclotron resonance mass spectrometers. But the principle of the invention can even be realized in inexpensive high frequency quadrupole ion trap mass spectrometers.

For longer genes, it is practical to process the DNA in segments of about 500 to 1000 base pairs in maximum. If no information is to be lost, the segments must overlap one another. For this purpose, at least partial sequences must therefore be known in order to produce the primers for the PCR. The segments are analyzed individually for mutations, each according to the same method as for short genes. The analysis of a candidate gene of a test subject proceeds in the same manner in a parallel analysis of segments. If the assignment of various alleles (mutants) of a gene to specific effects or symptoms of illness is known, it is not absolutely necessary to analyze the entire gene.

Using this method of application to many individuals, mutations can be found and their relative frequencies determined. The found mutations can be assigned to individual hereditary diseases by application of the method to groups of people with a known degree of relationship and with known hereditary diseases.

On the other hand, the gene of a test subject in which a malfunction of the corresponding protein is suspected can also be analyzed very easily for possible mutations using this method. A major advantage of this method is that practically all of the various mutation variants of a gene ("alleles") can be found which would lead to a clinically indistinguishable picture. The assignment of an individual to a population genotype which, for example, differs from others by a particular metabolic malfunction, can also be made simply in this way, since frequently several mutations or genetic variants of a single gene or, as a maximum, a few genes may be independently responsible for these various modes of functioning.

The creation of relatively short digest fragments by the restriction enzymes is not only particularly advantageous for the mass spectrometry involved but also for theoretical reasons. By short digest fragments, possibly appearing mutations can be separated from one another with a very high, calculable probability, so that practically only one mutation in maximum can be present in one digest fragment. Since two point mutations, as well as one insertion and one deletion each, can in unfortunate

situations compensate each other in mass such that they lead to exactly the same molecular weight, two such compensating mutations would no longer be recognizable. However, this case becomes extremely improbable with short digest fragments.

highest mass accuracy will be achieved Additionally, the above described method for polymorphism and mutation analysis offers redundant verification of the results, necessary for clinical applications. Substitution of a nucleotide within the coding ("sense") strand is automatically accompanied by a corresponding substitution within the complementary counterstrand ("antisense"). Since, even for double-stranded DNA samples in MALDI preparations, only single-stranded DNA is measured by MALDI-MS analysis, a shifting of molecular weights for both the sense as well as the antisense strand appears in mutated DNA digest fragments, thus corroborating the mutation.

Since new recognition sequence chains for the restriction endonucleases may also be produced due to DNA polymorphisms occurring, this can lead to the formation of newly occurring, smaller fragment masses during measurement. In addition, recognition sequence chains present on the wild type may be altered by a mutation and thus a sum fragment of higher mass including a base substitution could appear in place of the two fragment masses present for the wild type.

Additionally, the new method is capable of distinguishing between homo- and heterozygotic hereditary factors. Homozygotically polymorphous DNA sections demonstrate a complete shifting of two fragment masses as compared to the wild type in MALDI-TOF mass spectrometry. In heterozygosity, a spectrum occurs for a mutated and the wild type allele of a gene which demonstrates at least two additional fragment masses for sense and antisense in addition to the wild type spectrum, which corresponds to the polymorphous allele.

It is essential for the method proposed here that the digest fragments are purified and cleaned from all ions which may disturb the mass spectrometric measurement. For DNA fragments, adduct formation with sodium or potassium ions is detrimental for

mass determination. This is true for all kinds of ionization, ESI as well as MALDI. Effective cleaning methods have become known using magnetic beads with salt concentration-dependent surface adsorption, reverse-phase micro columns, or especially prepared pipette tips. Residual alkaline ions may be replaced by ammonium ions which are destroyed in the MALDI process.

The necessary chemicals and tools for PCR replication and for enzymatic digestion can be assembled in suitable kits and thus be used in a simple manner. This particularly applies to the analysis of mutations from DNA which can easily be determined by means of primers and corresponding sets of restriction enzymes. The tools for purification of PCR products and the digest products can also be included in the kits. The tools can, for example, consist of magnetic beads for temporary surface bonding for the purpose of washing, or also chromatographic mini-tubes or prepared pipette tips for purification. The kits may contain several enzymes, singly or in a few basic mixtures. For rapid genotyping, there may be mixtures for special genes or parts of genes, together with primers for the corresponding PCR.

In particular, the kits can be assembled and packed in such a way that they can be further processed by automatically operating pipetting robots.

The method need not be used exclusively for locating mutations in genes. With it, mutations in introns or regulatory DNA sections can also be studied and analyzed. The specialist in biogenetics is acquainted with further problems which can be easily solved using the principle of this invention.

Claims

1. A method for the mass spectrometric recognition of polymorphisms and mutations in a nucleic acid, comprising the following steps:
(1) providing amounts of double-stranded target segments of nucleic acid, preferably from one full or partial gene,
(2) adding a set of restriction enzymes, operating at similar buffer conditions at defined restriction points, to the DNA target segments, and digesting the target segments into a mixture of double-stranded DNA digest fragments of about 10 to 40 bases in length,
(3) removing from the mixture cations which may result in adduct formation by the ionization method chosen,
(4) determining the molecular weights of digest fragments of the mixture by mass spectrometry, and
(5) determining mutative changes or variations in the digest fragments by comparing the molecular weights of the digest fragments with those of a reference DNA, digested with the same set of endonucleases.

2. A method according to Claim 1, wherein the nucleic acid is a gene or gene segment.

3. A method according to Claim 1 or Claim 2, wherein the target segment of nucleic acid in step 2 is from one full or partial gene.

4. A method according to any one of the preceding claims, wherein the DNA target segments provided in step (1) are selected and amplified by PCR, if necessary after preliminary transcription of RNA into DNA.

5. A method according to any one of the preceding claims, wherein a selection of the restriction endonucleases for the set of endonucleases used in step (2) is carried out based on the known sequence of the DNA target segments and is performed to (a) generate digest fragment lengths of about 10 to 40 base pairs, (b) avoid overlap of the isotopic pattern of different digest fragment mass peaks,

and (c) use endonucleases of similar operation conditions and, if desired, low price.

6. A method according to Claim 5, wherein the said selection is carried out using a computer program.

7. A method according to any one of the preceding claims, wherein the MALDI process is used for ionization and a time-of-flight mass spectrometer.

8. A method according to Claim 7, wherein the spectrometer includes a reflector.

9. A method according to any one of the preceding claims, wherein one or more individual exons of a gene are simultaneously provided as DNA target segments.

10. A method according to any one of the preceding claims,. wherein larger exons are divided into overlapping DNA target segments that are subjected individually or together to mutation analysis.

11. A method according to any one of the preceding claims, wherein the masses of the usually overwhelming number of digest fragments, having the same mass as the corresponding digest fragment of the standard DNA, are used as internal mass reference peaks to improve mass determination of the digest fragments deviating in mass.

12. A method according to any one of the preceding claims, wherein the detection of mutations or variants in the resulting mass spectra in step (5) is automated by subtracting a reference DNA spectrum and observing the differences.

13. A method according to Claim 12, including the step of standardization.

14. A method according to any one of the preceding claims, wherein the reference DNA of step (5) is a standard or wild type DNA.

15. A method according to any one of Claims 1 to 12 for finding mutations in a diseased subject, wherein the reference DNA of step (5) is DNA of a healthy subject.

16. A kit for performing a mutation analysis for a given gene group, gene or partial gene according to one of the preceding claims, wherein the kit comprises at least the primers for PCR replication and a mixture of buffered restriction enzymes.

17. A kit according to Claim 16, which also contains other reaction components for PCR replication.

18. A kit according to Claim 17, which contains one or more of polymerases, activators and nucleotide triphosphates.

19. A kit according to any one of Claims 16 or 18, which also comprises one or more of magnetic beads, chromatographic micro-tubes or prepared pipette tips.

20. A kit according to any one of Claims 16 to 19, wherein all components are packed in such a way that they can be further processed by automatic pipetting robots.

21. A method for the mass spectrometric recognition of polymorphisms and mutations in a nucleic acid, substantially as described herein.

22. A kit for performing mutation analysis substantially as described herein.

| | | | |
|---|---|---|---|
| **Application No:** | GB 9914723.3 | **Examiner:** | Cass A. Compton Dottridge |
| **Claims searched:** | 1-22 | **Date of search:** | 22 November 1999 |

# Patents Act 1977
# Search Report under Section 17

**Databases searched:**

UK Patent Office collections, including GB, EP, WO & US patent specifications, in:

UK Cl (Ed.Q): G1B BAC

Int Cl (Ed.6): C12Q 1/68

Other: ONLINE: WPI, EPODOC, CAS-ONLINE.

**Documents considered to be relevant:**

| Category | Identity of document and relevant passage | Relevant to claims |
|---|---|---|
| X | WO98/20166 A2 (SEQUENOM) See whole document especially page 9, lines 25-27; page 15, line 20 - page 16, line 27 and example 12. | 1-8 and 12-16 |
| X | WO97/33000 A1 (GENETRACE SYSTEMS) See whole document especially page 17, line 32 - page 18, line 4; page 19, lines 18-24 and page 23, lines 27-30. | 1-8 and 12-16 |
| X | Rapid Communications in Mass Spectrometry **11**(15) 1997 "Detection of single nucleotide mutations [...]" Wada, Y. *et al.* pages 1657-1660. | 1-8 and 12-15 |
| X | Rapid Communications in Mass Spectrometry **11**(10) 1997 "Matrix-assisted laser desorption [...]" Jannavi, R-S. *et al.* pages 1144-1150. | 1-6, 8 and 12-15 |

| | | | |
|---|---|---|---|
| X | Document indicating lack of novelty or inventive step | A | Document indicating technological background and/or state of the art. |
| Y | Document indicating lack of inventive step if combined with one or more other documents of same category. | P | Document published on or after the declared priority date but before the filing date of this invention. |
| & | Member of the same patent family | E | Patent document published on or after, but with priority date earlier than, the filing date of this application. |